Structure
Ways & Means

An Approach for De Novo Structure Determination of Dynamic Molecular Assemblies by Electron Cryomicroscopy

Bjoern Sander,[1,2,*] Monika M. Golas,[2,3] Reinhard Lührmann,[4] and Holger Stark[2,5,*]
[1]Stereology and EM Laboratory, Institute of Clinical Medicine, Aarhus University, DK-8000 Århus C, Denmark
[2]Research Group of 3D Electron Cryomicroscopy, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany
[3]The Water and Salt Research Centre, Institute of Anatomy, Aarhus University, DK-8000 Århus C, Denmark
[4]Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany
[5]Göttingen Centre for Molecular Biology, University of Göttingen, D-37077 Göttingen, Germany
*Correspondence: bsander@ki.au.dk (B.S.), hstark1@gwdg.de (H.S.)
DOI 10.1016/j.str.2010.05.001

SUMMARY

Single-particle electron cryomicroscopy is a powerful method for three-dimensional (3D) structure determination of macromolecular assemblies. Here we address the challenge of determining a 3D structure in the absence of reference models. The 3D structures are determined by alignment and weighted averaging of densities obtained by native cryo random conical tilt (RCT) reconstructions including consideration of missing data. Our weighted averaging scheme (wRCT) offers advantages for potentially heterogeneous 3D densities of low signal-to-noise ratios. Sets of aligned RCT structures can also be analyzed by multivariate statistical analysis (MSA) to provide insights into snapshots of the assemblies. The approach is used to compute 3D structures of the *Escherichia coli* 70S ribosome and the human U4/U6.U5 tri-snRNP under vitrified unstained cryo conditions, and to visualize by 3D MSA the L7/L12 stalk of the 70S ribosome and states of tri-snRNP. The approach thus combines de novo 3D structure determination with an analysis of compositional and conformational heterogeneity.

INTRODUCTION

Molecular machines are at the heart of many central cellular processes such as activation, silencing, and repair of genes, transcription and processing of coding or noncoding RNAs, and cell signaling (Gavin et al., 2002; Nickell et al., 2006; Alberts, 1998). Single-particle electron cryomicroscopy (cryo-EM) occupies a central position in their structural investigation, as it often represents the only approach to make the macromolecule of interest directly visible. Reliable three-dimensional (3D) structures determined by cryo-EM are required in order to understand how architecture and dynamics of macromolecular complexes relate to their function. For many macromolecular assemblies, the amount of prior structural information is limited given their biochemical complexity (Wahl et al., 2009). Moreover, there are often unique technical challenges, both biochemically and structurally, that cannot be solved by a straightforward commonly applicable recipe. An important aspect impeding initial 3D structure determination is heterogeneity of macromolecular assemblies in terms of composition as well as conformation (Leschziner and Nogales, 2007). As for many transient macromolecular complexes, conformation and composition of a given complex may vary, being dependent on functional states (Mitra and Frank, 2006) or technical aspects such as purification procedure (Oeffinger et al., 2007).

Approaches for 3D structure determination in cryo-EM can be based on a priori or a posteriori assignment of projection angles, depending on the imaging technique. Angular reconstitution (AR) on the basis of the common lines principle (Crowther, 1971) is a technique for a posteriori determination of particles' projection angles. In a commonly used AR algorithm, sets of one-dimensional projections, so-called sinograms, are determined from two-dimensional particle views and correlated in an approach that identifies the common lines between projections taking into account the particles' symmetry (van Heel, 1987). This approach is regarded as a single-exposure technique, as reexposures under different specimen tilt angles are not necessary in order to determine the 3D structure. Common lines have also been applied to analyze particles with flexible domains (Hall et al., 2007). In alternative AR approaches, the particles' coordinate system is established by using, for example, asymmetric features of the particles as fiducial features in order to determine relative orientations (Rubinstein et al., 2003). Generally, AR does not provide information about the handedness of the structure and requires additional experiments to determine the handedness (Rosenthal and Henderson, 2003). Although AR can in principle yield high levels of resolution, the de novo 3D structure determination of asymmetric particles is often difficult, and the presence of structural heterogeneity in the data set along with a limited angular dispersion of the particle views can in practice hamper the identification of unique coordinates for the projections.

A priori information about the coordinate system to be established is available when particle images with known relative tilt angles are recorded using multiple-exposure techniques. For the random conical tilt (RCT) technique (Radermacher et al., 1987), one records two images per specimen area, one typically

40°–60° tilted as well as an untilted image, which establishes a partially defined coordinate system. When the untilted images show identical particle views only differing by their random in-plane rotation, this rotation together with the known tilt axis and tilt angle of the specimen can be used to fully describe a common 3D coordinate system for the particles. Classical RCT is typically applied to achieve a lowly resolved 3D reconstruction from one user-selected particle view including its handedness. This approach is considered reliable because it makes use of parameters predetermined in the electron microscope to assign particle orientations and thereby reduces computational complexity. However, as high image contrast facilitates the visualization of the tilt pairs and the RCT reconstruction, air-dried negatively stained specimens are typically used, which can lead to sample preparation artifacts (Cheng et al., 2006). Furthermore, standard RCT reconstructions are anisotropic because of missing data in directions not covered by the maximum specimen tilt. For samples with good angular dispersion on the specimen grid, the concept of RCT has thus been advanced to the orthogonal tilt (OT) reconstruction technique that uses two oppositely tilted (−45° and 45°) exposures per specimen area in order to increase the resulting relative tilt to 90°, which can yield isotropic 3D reconstructions (Leschziner and Nogales, 2006).

A complete a priori defined coordinate system can be obtained by electron cryotomography (cryo-ET) (Lučić et al., 2005). Cryo-ET differs from the other techniques mentioned here in its requirement to record many (~70) images of very low electron dose per specimen area, which in turn requires microscope hardware with the best possible characteristics in terms of beam and specimen stability as well as detector size and sensitivity. As for RCT, cryo-ET 3D reconstructions are anisotropic because of missing data in the regions not covered by the maximally possible tilt angle, which limits the attainable resolution.

Overall, a priori methods offer clear advantages in the absence of available reference 3D structures and the assumed presence of structural heterogeneity to establish an initial 3D structure of the sample. The general concept of 3D structure refinement in the presence of structural heterogeneity is to determine a set of 3D structures of the structural subgroups and to use these 3D models to competitively split up the image data set by supervised classification methods (Brink et al., 2004). This separation in silico represents a prerequisite to achieve better resolutions compared to merging the complete heterogeneous data set into a single 3D structure and to avoid artifacts such as the extinction of flexible domains in the final 3D structures (Penczek et al., 2006; Grob et al., 2006). Based on an available 3D structure representing the overall data set, structural heterogeneity was, for example, determined by resampling with replacement to yield data subsets which are clustered by 3D multivariate statistical analysis (MSA) and classification (Simonetti et al., 2008). Also, maximum-likelihood 3D classification (Scheres et al., 2009b), which likewise uses a low-resolution startup model, was applied to heterogeneous data sets in cases where high noise levels in the data prevented detection of heterogeneity by classical supervised classification methods.

The technical challenges associated with the two most common de novo reconstruction techniques, AR and RCT, require improvement of either AR or RCT. Here we decided to take advantage of the general simplicity and reliability of RCT reconstructions and extend the RCT protocol by (1) the possibility of using unstained cryo specimens and (2) the automated averaging of large sets of RCT reconstructions. This allowed us to increase the signal-to-noise ratio (SNR) of individual cryo-RCT maps and lowers the effects of the missing data in the individual reconstructions by merging individually rotated RCT maps. Thereby the presence of structural heterogeneity in many real-world data sets is accounted for by a weighted averaging strategy that works irrespective of whether the RCT reconstructions represent the same 3D structure (orientational diversity) or related, individual 3D structures (structural or mixed structural/orientational diversity). In the second step, MSA of the aligned RCT maps can be used to describe structural variability of the sample.

RESULTS

Concept of the Approach

The approach, in contrast to classical RCT, is not restricted to macromolecules that exhibit preferred particle binding, and it can thus deliver a better and more isotropic resolution. In contrast to conventional RCT, a smaller number of images can be used to compute individual RCT maps, as the SNR of the final structure increases by weighted averaging of the RCTs. This formation of small classes, in turn, facilitates an optimum separation of the particles in the 2D classification preceding the RCT reconstructions (Figure 1A). This approach thus leads to the generation of a large number of individual RCT 3D structures that are iteratively combined in a weighted average. By using concepts originating from work with noisy data (Sigworth, 1998) as well as from consideration of missing data in subtomogram averaging (Schmid and Booth, 2008), we achieved a high degree of independence of the starting point; that is, in all the various data sets analyzed by this approach so far, we could apply unbiased first references such as Gaussian noise.

This weighted RCT (wRCT) technique is suitable for unstained cryo images and can be followed by 3D MSA-based classification of the aligned RCT structures in order to detect structural heterogeneity in the sample. Therefore, we herein suggest taking advantage of the fact that the aligned RCT structures are directly suited for 3D MSA and 3D classification in order to detect structural subpopulations (Figure 1A). This second step of our approach might benefit the analysis of complex samples: compared to a data set comprising only 2D views with their three rotational and two translational degrees of freedom, RCT computation significantly reduces these degrees in the data set, as the particles within each RCT structure keep their fixed relative orientations. The reduction in degrees of freedom and—compared to individual particle images—the higher statistical reliability because of the pooling of signal within the RCTs are two unique, important factors preventing initial incorrectness of the alignment. Three-dimensional MSA and 3D classification are thus suggested to unravel heterogeneity by means of minimizing intraclass variance while maximizing interclass variance.

Reconstructions by RCT

Tilted image pairs (specimen-holder positions: first −45° and then 0°) of the assemblies were recorded and the untilted images were clustered by MSA-based unsupervised classification (van

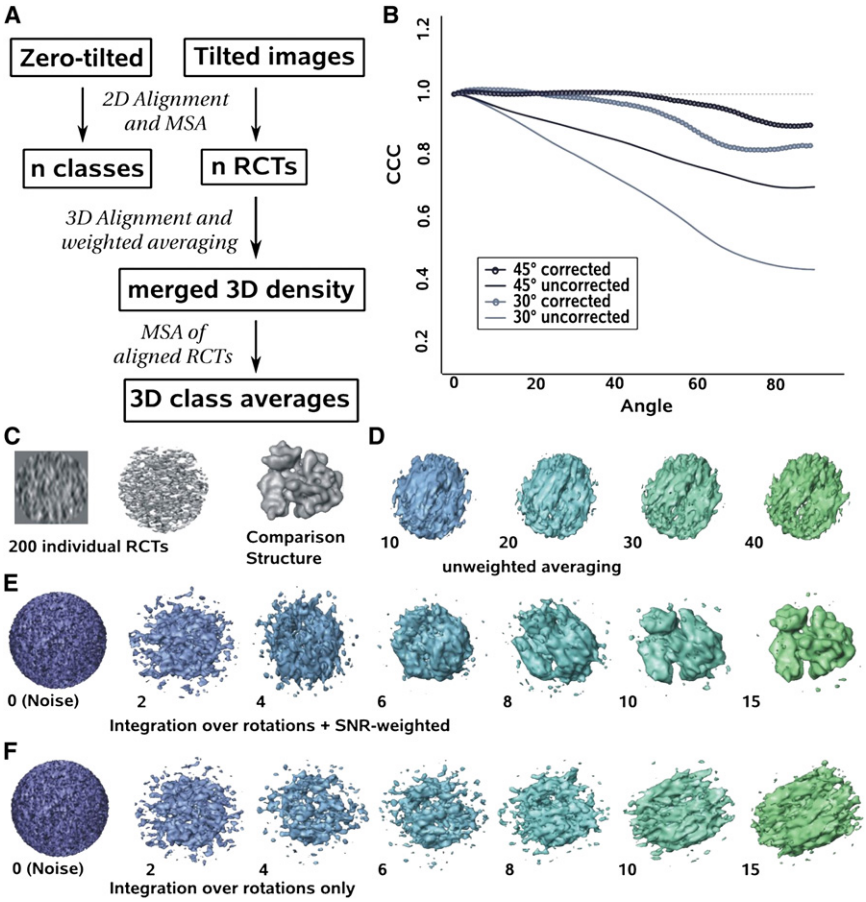

Figure 1. The wRCT Approach

(A) Flow chart summarizing the analysis approach used in this study.

(B) Cross-correlation between two RCTs of an identical ribosome structure as a function of missing data with and without the presence of missing data taken into account. The test has been carried out for specimen tilt angles of 45° and 30°, respectively. Without correction (continuous lines), the cross-correlation becomes reduced with increasing angular difference between the directions of the RCTs. With correction (bulleted lines), the effect can be compensated for to a good extent. CCC, cross-correlation coefficient.

(C) One of 200 individual RCTs computed from 41 ± 20 noisy images with high noise level and assumed specimen tilt angle of 27.5° shown as central section (left) and surface view (middle). The noise- and missing data-free structure is shown for comparison (right).

(D) Three-dimensional alignment and unweighted averaging starting with Gaussian noise as a first reference—even over 40 iterations—do not recover the structure.

(E) Application of our weighting scheme including integration over rotations and consideration of individual SNR recovers the structure over the course of 15 iterations.

(F) The same iteration scheme as in (E), but without considering the individual SNRs of the input RCTs does not recover the structure.

Heel, 1984). The tilted counterparts of each class are subjected to 3D reconstruction by classical RCT (Radermacher et al., 1987), except that unstained cryo images were used instead of the standard negatively stained particles. Therefore, the number of untilted images per class is equal to the number of tilted images per RCT. We used high numbers of images per class (~300) comparable to standard RCT, and also lowered the number of images per class down to ~40. Low numbers of images per class are suggested to minimize misclassification and mixing effects in the 2D classification and lead to the generation of many low-SNR RCT reconstructions. Thus, a typical data set consisting of several thousands of tilt pairs can generate many hundreds of individual, noisy RCT 3D structures. Generally, low-quality classes as indicated by, for example, lack of structural details or technical artifacts in the untilted class averages can be excluded from the RCT reconstruction.

Three-Dimensional Alignment

The alignment of the noisy 3D structures is implemented as rotational 3D alignment of the individual RCT reconstructions to a 3D reference structure. All RCT 3D densities are rotated against the reference density using an angular search grid with evenly spaced longitude and latitude angle pairs (Euler angles β and γ). The alignment algorithm is an extension of the alignment by resampling to polar coordinates previously applied to 2D images (Joyeux and Penczek, 2002). At each point of the search grid, a resampling of the 3D sections to polar coordinates takes place, and polar cross-correlation (CC) functions between target and reference sections are computed. In comparison to 2D polar alignments where only one radial averaging step of the polar CC functions is required in order to perform a one-dimensional rotational peak search, the 3D case requires two averaging steps of the stack of polar CC functions. The first step takes place in the z direction of the 3D volume and the second in the radial direction. This averaging yields one-dimensional CC functions that are used to find the maximum CC coefficient for rotating the input 3D density around the z axis (Euler angle α) in order to determine the optimum z axis rotation similar to in-plane polar alignments of 2D images. We also use the CC functions to extract and store the CC coefficients in order to be able to perform a CC-based weighted integration over all rotations afterward.

All measured CC coefficients are corrected for missing data effects. To this end, binary missing cone volumes (see Figure S1A available online) are computed in all directions of the search grid to model the data abundance in the RCTs. These missing cone volumes corresponding to each RCT are combined by weighted averaging in parallel to the weighted averaging of the RCTs. Thereby, any cone's overlap with the weighted average of all cone volumes from the last iteration is determined. This estimate for the Fourier overlap of RCT and reference 3D structure is then

used for correction by dividing the measured CC coefficients by the actual Fourier overlap (Schmid and Booth, 2008). For all orientations tested during 3D alignment, the CC coefficients, inner products, and root mean squares are stored for weighted integration over all rotations in the subsequent step. For the first round of 3D alignment, an arbitrary unbiased first reference such as pure Gaussian noise is applied. In all subsequent rounds of the iterative alignment and weighted averaging cycle, the respective last weighted average is selected as reference.

Analysis of Fourier Overlap

In general, the CC between two zero-averaged complex vectors a and b of length n is

$$CC(a,b)=\frac{\sum_{i=1}^{n}(Re(a_i)\cdot Re(b_i))+\sum_{i=1}^{n}(Im(a_i)\cdot Im(b_i))}{\sqrt{\sum_{i=1}^{n}\left(Re(a_i)^2+Im(a_i)^2\right)\cdot\sum_{i=1}^{n}\left(Re(b_i)^2+Im(b_i)^2\right)}}. \quad (1)$$

When we denote the real and imaginary parts of a, $x_i = Re(a_i)$ and $y_i = Im(a_i)$, respectively, then the approximations $x_{1..n}\approx x_0$ and $y_{1..n}\approx y_0$ turn out to be sufficient to estimate the overlaps of, for example, RCT structures and weighted RCT averages. From Equation 1, one can also deduce an autocorrelation $CC(r,r')$, where r and r' are two representations of an identical object, each with individual regions of missing data. With the above-mentioned approximation, the overlap $O_{p,q}$ between binary cones or weighted averages of such cones, p and q, that have been rotated together with r and r', can then be estimated as

$$O_{p,q}\approx\frac{2\sum_{i=1}^{n}p_iq_i}{\sqrt{\sum_{i=1}^{n}2p_i^2\sum_{i=1}^{n}2q_i^2}}. \quad (2)$$

The accuracy of this approximation is dependent on similar statistical properties of the underlying 3D structure in missing versus nonmissing regions. We have therefore tested Equation 2 in simulation experiments with Gaussian-distributed vectors of different lengths and data absence and found that the error was less than 1% for $n=16^3$ and less than 0.1% for $n=96^3$, which is sufficiently low even for small 3D volumes. In practice, we see that inaccuracy can also originate from "leakage" of information into the missing cone that is generated by the ability of the 3D reconstruction algorithm to introduce some 3D information into Fourier regions outside those covered by the central section theorem. Based on our tests, we have found that 3D reconstructions using the weighted backprojection algorithm of IMAGIC-5 (van Heel et al., 1996) show added information equivalent to an additional specimen tilt of ~10°, and we have considered this in the computations. The dependence of the CC coefficients on the rotation of the missing cone for 45°- and for 30°-tilted specimens as well as the effectiveness of our correction approach to restore the CC coefficient heights to a very good extent can be seen in Figure 1B. The effect of the correction on the iterative alignment described below is characterized by a faster convergence to the underlying structure and less susceptibility to undesired smearing artifacts caused by the missing cones (Figure S1B).

Three-Dimensional Weighted Averaging

For the weighted integration over all rotations, we use exponentials of the CC coefficients as weights. In Sigworth (1998), a model Θ with $\Theta=(A,\sigma)$ was defined which consists of the underlying structure A and the standard deviation of the white Gaussian noise σ. This allowed for estimation of the unknown parameters A and σ based on known observations of the model, that is, correlation coefficients of the alignment with the resulting weighting factors being exponential transformations of the correlation coefficients. When applied to RCT densities, the advantage of this weighting is that it enhances input 3D densities with unique, high CC peaks while it reduces the influence of inferior structures exhibiting low correlation coefficients over all alignment parameters. When we apply this model to the 3D densities, every input density X_i is considered as a representation of the underlying 3D structure A, each randomly rotated by the three Euler angles α, β, and γ defined as in van Heel (1987) with applied independent noise R_i:

$$X_i=A(-(\alpha_i,\beta_i,\gamma_i))+\sigma R_i. \quad (3)$$

When we additionally consider an individual signal contribution σ_s of A in the 3D densities as well as data normalization during preprocessing, Equation 3 becomes

$$X_i=\frac{\sigma_{s,i}\,A(-(\alpha_i,\beta_i,\gamma_i))+\sigma_iR_i}{\sqrt{\sigma_{s,i}^2+\sigma_i^2}}. \quad (4)$$

In Sigworth (1998), the formulation of a *log* probability L that a particular data set X arises from the model with parameter set Θ led to an iterative scheme with integration over all alignment parameters:

$$A^{n+1}=\frac{1}{N}\sum_{i=1}^{N}\frac{\int X_i(\varphi)\gamma_i(\varphi;\Theta^{(n)})d\varphi}{\int\gamma_i(\varphi;\Theta^{(n)})d\varphi} \quad (5)$$

with a re-estimation of the standard deviation of the noise given as

$$\sigma^{n+1}=\left(\frac{1}{NM}\sum_{i=1}^{N}\frac{\int|X_i(\varphi)-A^{(n)}|^2\gamma_i(\varphi;\Theta^{(n)})d\varphi}{\int\gamma_i(\varphi;\Theta^{(n)})d\varphi}\right)^{\frac{1}{2}}, \quad (6)$$

where γ_i is the exponential of the (unnormalized) correlation coefficient divided by the squared standard deviation of the noise.

To account for individual SNR_i as in Equation 4, a weighting of each RCT structure X_i by multiplication with a weighting factor w_i can be performed. One obvious source of such individual SNR_i of the input 3D densities originates from the individual numbers of images per RCT, that is, an RCT computed from a small cluster exhibits a lower SNR as an RCT of a large cluster. Coming from Equations 5 and 6, the weighted average A and model parameter σ can be extended:

$$A_w^{n+1}=\frac{1}{N}\sum_{i=1}^{N}w_i\frac{\int X_i(\varphi)\gamma_i(\varphi;\Theta^{(n)})d\varphi}{\int\gamma_i(\varphi;\Theta^{(n)})d\varphi} \quad (7)$$

$$\sigma_w^{n+1}=\left(\frac{1}{NM}\sum_{i=1}^{N}w_i\frac{\int|X_i(\varphi)-A^{(n)}|^2\gamma_i(\varphi;\Theta^{(n)})d\varphi}{\int\gamma_i(\varphi;\Theta^{(n)})d\varphi}\right)^{\frac{1}{2}}. \quad (8)$$

When factors w_i are selected that represent an estimate of the SNR_i of the input 3D densities, this enhances RCT structures that exhibit a good SNR and thereby supports convergence of the iterative scheme. Here, SNRs are deduced from the maximum CC coefficient ccc_{max} (Frank and Al-Ali, 1975) of each input 3D density:

$$SNR_i \approx \frac{ccc_{max,i}}{1 - ccc_{max,i}}. \quad (9)$$

We compute weighting factors w_i that are roughly proportional to the SNR_i from Equation 9, but take additional account of the fact that the input RCT structures have been normalized to identical standard deviations before iterative alignment without knowledge of their individual SNR_i (see Supplemental Experimental Procedures). We repeat this scheme consisting of 3D alignment and weighted averaging iteratively typically 15–40 times using a progressively decreasing search grid with 15° separation in the first iteration and 2° in the final one. In addition to fixed, evenly spaced search grids, we have also included the option to generate search grids from random numbers and noticed that the convergence behavior of our data sets is practically independent of the exact angular distribution.

Synthetic Performance Tests

To analyze this approach, we generated synthetic RCT data and subjected them to iterative alignments. As seen in Figure 1C, we generated 200 RCTs from noisy projections assuming a cluster size distribution of 41 ± 20 images per RCT (for a representative RCT reconstruction, see Figure 1C). The RCT volumes had a strong missing cone assuming a specimen tilt angle of only 27.5°. Such a low tilt angle and resulting large extent of missing data can be seen as a stress test for the alignment of RCTs, as in practice specimen tilt angles of >40° are used for acquisition of the tilt pairs. Iterative alignment and averaging of the 200 volumes started with pure Gaussian noise as initial reference. When the averaging was performed without weighting (Figure 1D), no convergence is seen even for long schemes with 40 iterations, as the averages rapidly adopt a missing cone smear from which the system does not recover. When we performed an integration according to Equations 7 and 8 (integration over rotations and additional SNR weighting of individual RCTs; Figure 1E), this leads to an isotropically defined 3D average structure with strongly enhanced signal over the course of 15 iterations, whereas the integration over rotations only (Equations 5 and 6) (Figure 1F) again leads to smearing artifacts caused by the alignment of missing cones.

When the synthetic RCT reconstructions were recomputed using either higher SNRs, a higher specimen tilt angle, or a combination of both, unweighted averaging was sufficient to ensure convergence to the correct underlying structure very similar to the case shown in Figure 1E. Importantly, in all cases of convergence, the resulting average 3D structure was practically free of "missing cone" artifacts as a result of averaging RCTs with different rotation angles. Thus, our wRCT approach—in contrast to classical RCT—profits from angular diversity of the molecular images. In this context, it should be mentioned that any correct averaging of RCTs originating from different individual rotational angles of the particle on the specimen grid leads to an improvement in terms of isotropy and resolution of the average compared to an individual RCT, even when an isotropic particle distribution on the specimen grid cannot be achieved technically. We therefore recommend our approach irrespective of the specimen's adsorption characteristics. The CPU time needed to carry out such an iterative cycle of 15 rounds on 200 RCTs of 64 × 64 × 64 voxels was 3.5 hr on an eight-core Xeon X5550 system with 2.66 GHz (Intel), so computational power is not an issue with our approach.

Analysis of Ribosomes and Tri-snRNP Particles by wRCT

The characteristics of the two data sets are summarized in Table S1. The structure determination of the 70S *E. coli* ribosome was based on 27,748 image pairs in vitrified native buffer (Figure S2). The untilted images were clustered into 694 classes (i.e., ~40 images per class). A set of 489 class averages was selected for further RCT reconstruction leading to 489 RCT 3D structures computed from the tilted counterparts of the particle images. These RCT reconstructions were subjected to 40 rounds of iterative alignment and weighted averaging according to Equations 7 and 8 with gradually finer search grids starting from 15° separation to 2° in the final round. We have tested the robustness of the approach by using incorrect 3D models (Gaussian noise, two-fold symmetrized ribosome, ribosome of wrong handedness, tri-snRNP 3D structure) as starting references for the alignment. The correct structure was obtained for all the different initial 3D references tested (Figure 2A). Analyzing 10,177 tilted image pairs of human tri-snRNPs purified as in Sander et al. (2006) and recorded under unstained, vitrified cryo conditions (Figure S2), the consensus 3D structure was determined using the same iteration scheme as for the ribosome. The number of classes for the classification of the untilted images was 254 (i.e., ~40 images per class), and the RCT data set comprised the 254 individual RCTs. As for the ribosome, using different starting points including wrong structures did not influence the final 3D result for the tri-snRNP (Figure 2B). The final 3D map obtained is in excellent agreement with our previously published one (Sander et al., 2006) (Figure 2C). Also, the final ribosome structure, when evaluated by rigid-body fitting to crystallographic data of the *E. coli* ribosome (Schuwirth et al., 2005), revealed excellent agreement in overall shape and architectural details (Figure 2D). Overall, the structures became visible typically within the first ten iterations (Figures 2E and 2F), with the following iterations being mainly refinement rounds.

Adsorption Characteristics and Resolution

The angular plots of the two data sets (Figures 2G and 2H) show the Euler angles belonging to the direction of the maximum CC together with the weight w_i (Equations 7 and 8) assigned to this density and represent the typical adsorption characteristics of the two particles. Fourier-shell-correlation (FSC) curves were calculated upon splitting of the RCT members into two halves while preserving their original weighting; they indicated ~26–29 Å resolution for the ribosome and ~34–47 Å for the tri-snRNP using the 3σ and $FSC_{0.5}$ criteria (Figure 2I). The $R_{measure}$ (Sousa and Grigorieff, 2007) $FSC_{0.5}$ resolution was ~30 Å for the tri-snRNP and ~31 Å for the ribosome. A contrast transfer function (CTF) correction of the tilted particle images (Sander et al., 2003a) indicated a defocus of 7.5 μm ± ~1.3 μm for the tri-snRNP and of ~2.3 μm ± 0.8 μm for the ribosome (Figures S2E–S2H).

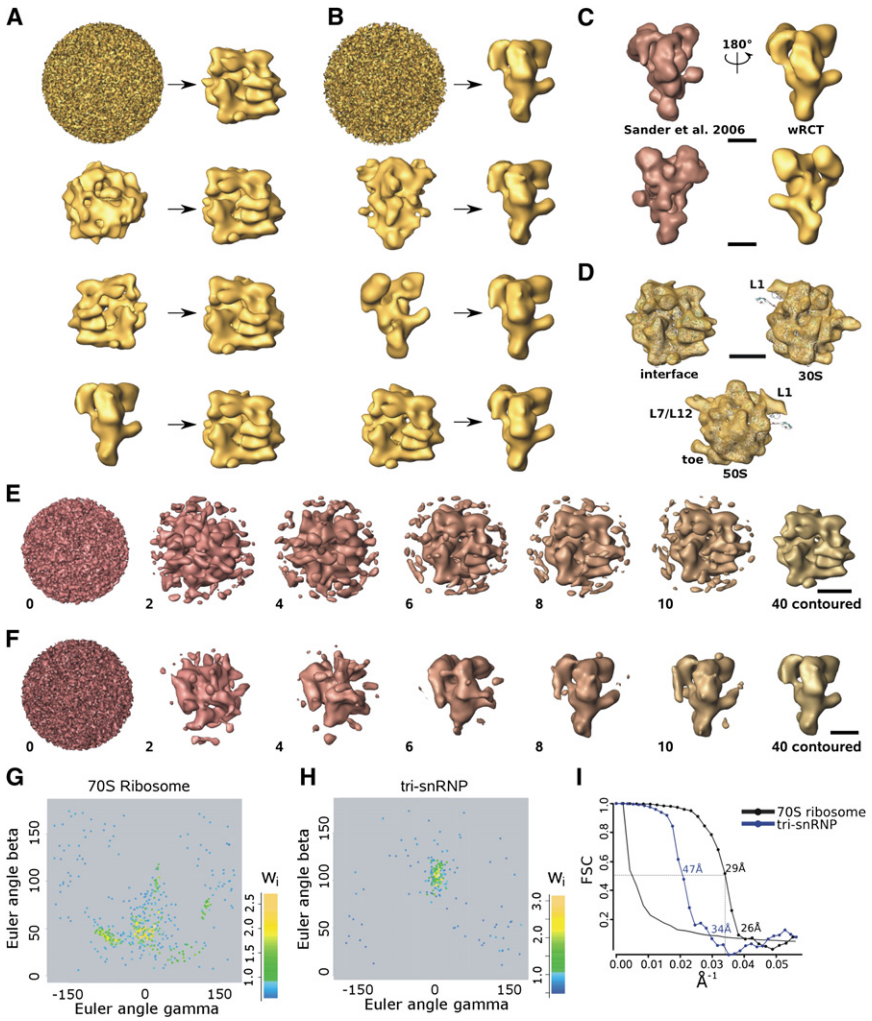

Figure 2. Cryo-EM Data Computations

(A) Convergence of 489 70S ribosome RCT structures computed from unstained, vitrified tilt pairs (~40 images per RCT) to the ribosome 3D structure using various incorrect initial references (top to bottom: Gaussian noise, C2-symmetrized ribosome map, ribosome of wrong handedness, tri-snRNP).

(B) Convergence of 254 tri-snRNP RCT structures computed from unstained, vitrified tilt pairs to the tri-snRNP 3D structure using various incorrect initial references (top to bottom: Gaussian noise, C2-symmetrized tri-snRNP map, tri-snRNP of wrong handedness, ribosome).

(C) Final wRCT 3D structure of the tri-snRNP compared to the previously published structure.

(D) Final wRCT 3D structure of the ribosome with fitted crystallographic coordinates (PDB ID codes 2AW7 and 2AWB) visualized at the interface site, the 30S subunit, and the 50S subunit.

(E and F) Convergence of the ribosome (E) and the tri-snRNP (F) over the course of 40 iterations.

(G and H) Euler angle plot with color-coded weights w_i of the RCTs of the ribosome (G) and of the tri-snRNP (H).

(I) Fourier-shell-correlation (FSC) analysis indicates a resolution at FSC = 0.5 of ~29 Å (ribosome) and ~47 Å (tri-snRNP). At the 3σ threshold, the resolution was ~26 Å (ribosome) and ~34 Å (tri-snRNP).

The scale bars represent 10 nm.

The resolution of the final wRCT averages approached the first CTF zero of the low-to-medium defocused images irrespective of the CTF correction, and the overall appearance of the structures originating from the CTF-corrected and -uncorrected images was very similar (Figures S3A, S3B, S3D, and S3E).

Performance of wRCT Using RCTs of Large Classes

Using an average cluster size of ~300 images in the classification of the untilted images, the 27,748 zero-tilt images of the 70S ribosome yielded 92 classes. This partition was used to compute 92 RCT 3D structures from the tilted counterparts of the tilt pairs. Using such a high number of images per class, one could expect that the classes led to a good signal in the RCT reconstruction that is sufficiently high for isosurface rendering. In total, ~30 RCT structures out of the 92 were of sufficient quality so that the ribosome was recognizable in the isosurface (Figure 3A, upper row). For the majority of classes, the structures showed noisy, partially incomplete structures despite the presence of ~300 images in the structure (Figure 3A, middle and bottom rows). The quality of the reconstruction did not correlate with the number of images per class, and it was also not possible to deduce the RCT quality visually from the corresponding zero-tilt class averages (Figure 3B). Nevertheless, the wRCT converged to the underlying structure including downweighting of RCT structures of low quality. Upon wRCT, the distribution of correlation coefficients of the individual RCT structures to the final wRCT average reflected the higher number of images per RCT (Figure 3C). For example, for the classification using ~300 images per class, 60 RCT structures showed a maximum correlation higher than 0.3; for the classification using ~40 images per class, 138 RCT structures showed a correlation higher than 0.3. Importantly, independent of the number of classes, the final weighted average upon wRCT was better defined (Figure 3D) and showed a higher isotropy (Figure 3E) than any of the individual RCT reconstructions. The resolution and overall appearance of the wRCT average based on ~300 images per class was very similar to the structure based on the classification with ~40 images per class described above (compare also Figures S3C–S3E).

Three-Dimensional MSA

After the alignment of the large number of RCT structures, individual RCT structures may still be representative of the variety of molecules that differ in composition or conformation. This heterogeneity within the sample may be identified within the set of RCT densities by clustering based on MSA (Liu et al., 2004) and hierarchical ascendant classification (HAC) (van Heel, 1984) on matrices of the complete set of 3D densities. In the present study, 3D classes comprising on average of five to eight

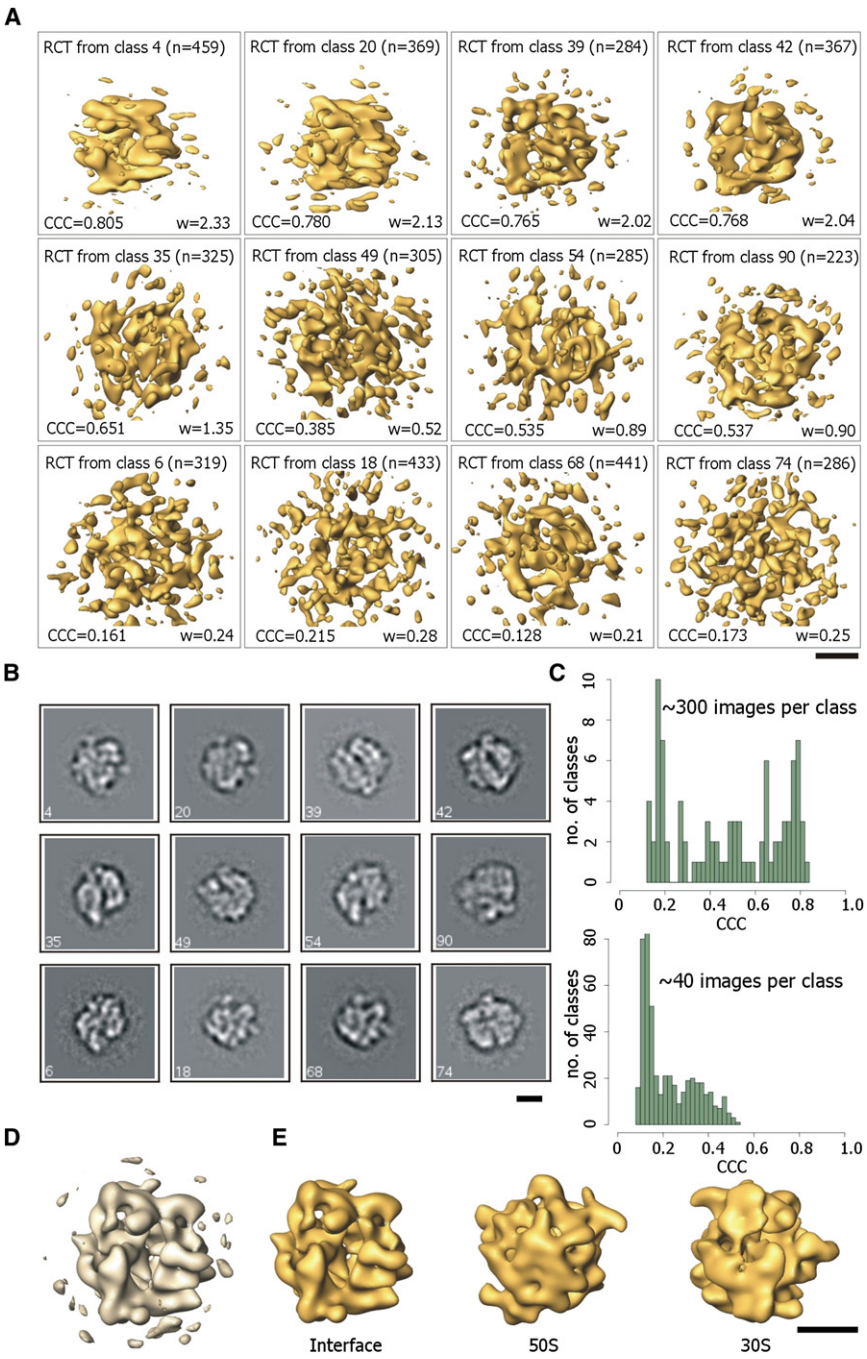

Figure 3. The wRCT Structure Determination of 70S Ribosome Based on Classification with Approximately 300 Images per Class

(A) The classification of the 27,748 zero-tilt images yielded 92 classes. This partition was used to compute 92 RCT 3D structures from the tilted counterparts. Uncontoured isosurface representations of representative RCT 3D structures are shown here along with the class number, the number *n* of images within the class, the maximum cross-correlation coefficient (CCC) to the final wRCT average, and the resulting weight *w* within the final wRCT average. In the first row, 3D structures with high correlation are shown; in the second row, 3D structures with intermediate correlation; and in the third row, 3D structures with low correlation. In total, ~30 RCT structures out of the 92 had a sufficient quality so that the ribosome was recognizable in an isosurface rendering without further averaging.

(B) Zero-tilt class averages of the respective RCT structures shown in (A).

(C) Distribution of maximum correlation coefficients of the individual RCT structures to the final wRCT average in comparison to the classification using ~40 images per class. For the classification using ~300 images per class (top), 60 RCT structures showed a correlation higher than 0.3; for the classification using ~40 images per class (bottom), 138 RCT structures showed a correlation higher than 0.3.

(D) Uncontoured isosurface rendering of the final wRCT ribosome structure at the interface site.

(E) Three views of the final wRCT structure (contoured; the ribosome is shown at the interface side, the 50S side, and the 30S side).

The scale bars represent 10 nm.

RCT structures corresponding to typically ~200–400 particle images are formed (Figure 4). We suggest that averaging of only small numbers of RCT structures avoids averaging of molecules that differ in composition or conformation.

The search for variable domains of U4/U6.U5 tri-snRNP by this 3D MSA of the aligned RCT maps indicates differences in three main domains, particularly a stalk domain on the right side of the main body (Figure 4A), the left head domain belonging to the U5 subunit of tri-snRNP (Figure 4B), and the U5 foot domain (Figure 4C). Previously, the variability of the U5 head domain was described in negatively stained images (Sander et al., 2006).

Besides these three prominent sources of structural diversity, the 3D class averages also indicate that additional smaller foci of diversity exist in the upper part of the particle that become characterizable once increased resolution of the technique is accompanied by additional structural information. In certain 3D MSA subclasses of the ribosome, density located at the C-terminal domains of the L7/L12 stalk is visualized (Figures 4D and 4E). Owing to their mobility, density for this part of L7/L12 usually escapes detection by X-ray crystallography or cryo-EM.

DISCUSSION

The objective of this work is to develop the capabilities of single-particle EM analysis with respect to determining a 3D structure from scratch when no suitable low-resolution reference models exist to start the computations directly with the alignment of 2D projection images. Whereas a set of 2D class averages can often be obtained quickly, the determination of the first 3D structure

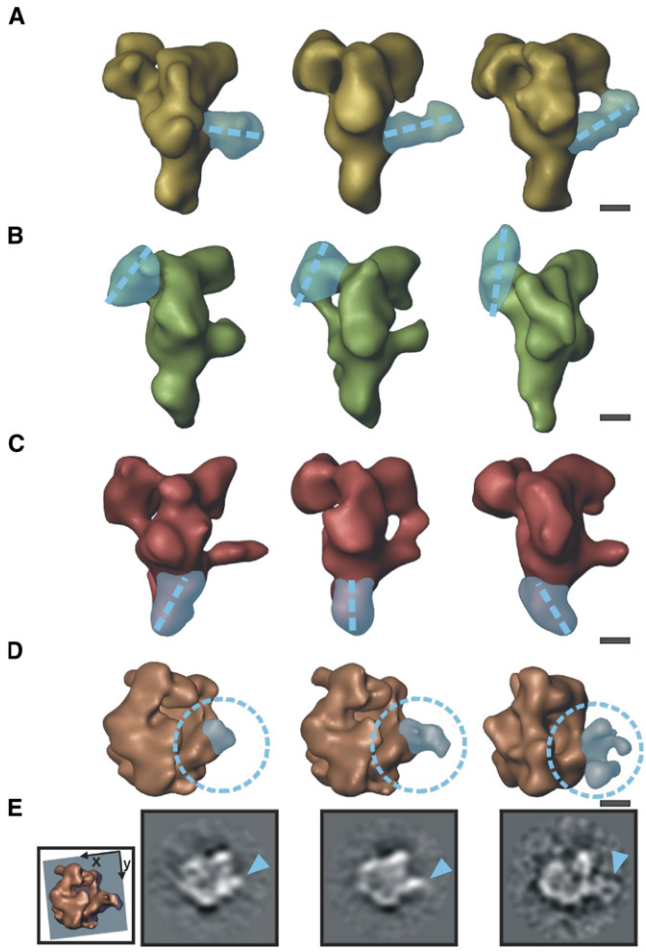

Figure 4. Analysis by 3D MSA

(A) Three selected 3D class averages of human U4/U6.U5 tri-snRNP which differ in orientation of the stalk (blue).

(B) Three selected 3D class averages of tri-snRNP with an individual main axis of the left U5 head domain (blue).

(C) Three selected 3D class averages of tri-snRNP with a deviation in foot orientation (blue).

(D and E) Three selected 3D class averages of 70S *E. coli* ribosome with no visible L7/L12 density (left), partly visible stalk (center), and fully visible stalk (right). For comparison, central sections of the 3D class average through the stalk are shown in (E) (orientation of the section as depicted in the inlay on the left; blue arrowheads point toward the stalk density). The C-terminal domains of L7/L12 are functionally important for factor recruitment and GTPase activation. Density that can accommodate up to four C-terminal domains can be seen as separate density elements, in line with biochemical data (Diaconu et al., 2005).

The scale bars represent 5 nm.

representing these averages can be a laborious and unpredictably lengthy process. This challenge becomes more critical the less "well behaved" the sample is. The effort necessary to determine a first reliable structure of an as yet structurally uncharacterized complex is dependent on multiple factors, such as its biochemical susceptibility to degradation, the presence of flexible domains, undesired copurification of biochemically similar particles, or biochemical limitations in order to achieve a functionally defined state of a macromolecule that undergoes structural changes. Under such adverse situations, it is difficult to decide whether the individual projection views that can be detected in a macromolecular sample represent views belonging to the same 3D object or to different heterogeneous models. At low resolutions, the answer may often be "yes and no," as the preliminary approximation that most or all views belong to the same 3D model can be acceptable in the beginning, but has to be given up when the resolution of the map is to be refined. The approach presented here was shown to be useful to generate such a first consensus model, and it has also turned out to be applicable even for very heterogeneous specimens (Golas et al., 2009) where the combined lack of knowledge about a particle's structure, individual deviations from the overall consensus model due to heterogeneity, as well as the particle's adsorption characteristics were unknown parameters to be determined at the same time. In this context, the classical RCT technique suffers from the fact that a single RCT 3D structure is only one possible representative of a dynamic macromolecule. In contrast, the automated integration of all data into a consensus 3D model described here, followed by 3D MSA and classification, is reliable, user independent, and compatible with the pronounced compositional and conformational dynamics of many macromolecular complexes. Thus, this approach can provide the initial 3D structural information needed in order to proceed with a further analysis of the sample on the basis of classical (Penczek et al., 1994) or maximum-likelihood-driven (Scheres et al., 2007, 2009a) projection-matching and classification techniques that benefit from initial 3D models. As untilted electron microscopic images often show a considerably higher resolution than tilted images, it is thus possible to improve the resolution of an initial wRCT 3D structure by using this structure as a reference to refine the alignment of untilted particle images.

The resolution of the wRCT 3D structures presented here approaches the first zero of the CTF function corresponding to the low-to-intermediate defoci present in the tilted data set irrespective of a CTF correction, indicating that the high B factor and limited resolution of the tilted cryo images limit the resolution. We therefore suggest that improvements attainable on the hardware side, such as elimination of focus or magnification gradients (Zemlin, 1989), use of tilt-stable cryo-specimen boxes at high specimen tilt angles, and better detectors (McMullan et al., 2009) will also benefit the resolution of wRCT structures. The technique presented here is in principle expandable to cryotomography data, and both cryo-RCT and -tomography techniques will take advantage of current and future EM hardware developments. Likewise, our approach can be applicable to 3D densities from orthogonal tilt (OT) reconstruction techniques (Leschziner and Nogales, 2006) that eliminate missing data regions with their 45°/−45° tilt pairs. However, many samples show preferential angular adsorption properties that are difficult to handle by OT, so it should be noted that transitions between the RCT and OT recording schemes with tilt angle pairs individually optimized to a sample's adsorption characteristics also present a good option. Overall, we thus suggest that especially structural research on highly dynamic assemblies will benefit from these methodological advances. The performance of our approach on dynamic particles strongly suggests a general applicability for other complex macromolecular machines that currently are only poorly understood.

EXPERIMENTAL PROCEDURES

Electron Microscopy

Particles were adsorbed onto carbon film and plunge-frozen into liquid ethane for imaging in vitrified buffer (Adrian et al., 1984). Images were taken on a CM 200 FEG electron microscope (FEI) at 160 kV. A 4 k × 4 k CCD camera (TemCam-F415; TVIPS) was used at 2-fold binning of the pixels and 122 k× magnification.

Image Processing

Untilted particle images were subjected to five to ten rounds of alignment by using polar coordinates (Sander et al., 2003b), followed by MSA and 2D class averaging. RCT structures were computed for each class (Radermacher et al., 1987) by exact filtered backprojection with refinement of orientation of the tilted images by translational parallel alignment against low-pass-filtered reprojection images within IMAGIC-5 (van Heel et al., 1996). Before 3D alignment of 3D RCT volumes, the RCT structures were band-pass filtered. The pixel sizes were ~5 Å/pixel (ribosome) and ~7 Å/pixel (tri-snRNP).

Three-Dimensional MSA

After 40 iterations of 3D alignment and weighted averaging, the 3D volumes were appended into a single file in the orientation belonging to the maximum CC and treated by MSA followed by HAC and class averaging within the IMAGIC-5 software. The average number of RCT volumes contributing to one 3D class average was five to eight.

Synthetic RCT Volumes

Generated 2D projections from an ideal 3D structure from Protein Data Bank (PDB) ID codes 1FFK and 1FJG (Ban et al., 2000; Carter et al., 2000) were adjusted to SNR = 0.005 by addition of noise before weighted backprojection. A constant tilt angle was assumed to generate the typical missing cone of RCT structures. The noisy density map was randomly rotated to simulate a dispersion of orientations.

Availability of Software

The 3D alignment and weighted averaging software is available upon request to bsander@ki.au.dk.

ACCESSION NUMBERS

The 3D maps of the 70S ribosome and the U4/U6.U5 tri-snRNP have been deposited in the EM Data Bank under accession numbers EMD-1728 and EMD-1729.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and one table and can be found with this article online at doi:10.1016/j.str.2010.05.001.

ACKNOWLEDGMENTS

We are grateful to F.J. Sigworth for comments on the manuscript. We are grateful to M. Rodnina for providing ribosomal samples. We thank T. Conrad, P. Kempkes, and H. Kohansal for expert technical assistance. The work was supported by grants from the German Federal Ministry of Education and Research (BMBF) and a European "3D Repertoire" grant (to H.S.) and the Ernst-Jung-Stiftung and "Fonds der Chemischen Industrie" (to R.L.). B.S. is grateful for support by the Novo Nordisk Foundation, the Agnes og Poul Friis' Foundation, Fru Astrid Thaysens Legat for Lægevidenskabelig Grundforskning, and the A.P. Møller Foundation for the Advancement of Medical Sciences. M.M.G. is supported by the Danish Council for Independent Research, Medical Sciences and the Agnes og Poul Friis' Foundation. B.S. and M.M.G are grateful for financial support by the Danish Center for Scientific Computing (DCSC) as well as for assistance by Niels Carl W. Hansen, DCSC/AU. The authors declare no competing financial interests.

Received: December 9, 2009
Revised: April 8, 2010
Accepted: May 4, 2010
Published: June 8, 2010

REFERENCES

Adrian, M., Dubochet, J., Lepault, J., and McDowall, A.W. (1984). Cryo-electron microscopy of viruses. Nature *308*, 32–36.

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell *92*, 291–294.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science *289*, 905–920.

Brink, J., Ludtke, S.J., Kong, Y., Wakil, S.J., Ma, J., and Chiu, W. (2004). Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. Structure *12*, 185–191.

Carter, A.P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Wimberly, B.T., and Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature *407*, 340–348.

Cheng, Y., Wolf, E., Larvie, M., Zak, O., Aisen, P., Grigorieff, N., Harrison, S.C., and Walz, T. (2006). Single particle reconstructions of the transferrin-transferrin receptor complex obtained with different specimen preparation techniques. J. Mol. Biol. *355*, 1048–1065.

Crowther, R.A. (1971). Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. Philos. Trans. R. Soc. Lond. B Biol. Sci. *261*, 221–230.

Diaconu, M., Kothe, U., Schlünzen, F., Fischer, N., Harms, J.M., Tonevitsky, A.G., Stark, H., Rodnina, M.V., and Wahl, M.C. (2005). Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. Cell *121*, 991–1004.

Frank, J., and Al-Ali, L. (1975). Signal-to-noise ratio of electron micrographs obtained by cross correlation. Nature *256*, 376–379.

Gavin, A., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A., Cruciat, C., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

Golas, M.M., Böhm, C., Sander, B., Effenberger, K., Brecht, M., Stark, H., and Göringer, H.U. (2009). Snapshots of the RNA editing machine in trypanosomes captured at different assembly stages in vivo. EMBO J. *28*, 766–778.

Grob, P., Cruse, M.J., Inouye, C., Peris, M., Penczek, P.A., Tjian, R., and Nogales, E. (2006). Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues. Structure *14*, 511–520.

Hall, R.J., Siridechadilok, B., and Nogales, E. (2007). Cross-correlation of common lines: a novel approach for single-particle reconstruction of a structure containing a flexible domain. J. Struct. Biol. *159*, 474–482.

Joyeux, L., and Penczek, P.A. (2002). Efficiency of 2D alignment methods. Ultramicroscopy *92*, 33–46.

Leschziner, A.E., and Nogales, E. (2006). The orthogonal tilt reconstruction method: an approach to generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles. J. Struct. Biol. *153*, 284–299.

Leschziner, A.E., and Nogales, E. (2007). Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions. Annu. Rev. Biophys. Biomol. Struct. *36*, 43–62.

Liu, J., Reedy, M.C., Goldman, Y.E., Franzini-Armstrong, C., Sasaki, H., Tregear, R.T., Lucaveche, C., Winkler, H., Baumann, B.A.J., Squire, J.M., et al. (2004). Electron tomography of fast frozen, stretched rigor fibers reveals elastic distortions in the myosin crossbridges. J. Struct. Biol. *147*, 268–282.

Lučić, V., Förster, F., and Baumeister, W. (2005). Structural studies by electron tomography: from cells to molecules. Annu. Rev. Biochem. *74*, 833–865.

McMullan, G., Faruqi, A.R., Henderson, R., Guerrini, N., Turchetta, R., Jacobs, A., and van Hoften, G. (2009). Experimental observation of the improvement in MTF from backthinning a CMOS direct electron detector. Ultramicroscopy *109*, 1144–1147.

Mitra, K., and Frank, J. (2006). Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. Annu. Rev. Biophys. Biomol. Struct. *35*, 299–317.

Nickell, S., Kofler, C., Leis, A.P., and Baumeister, W. (2006). A visual approach to proteomics. Nat. Rev. Mol. Cell Biol. *7*, 225–230.

Oeffinger, M., Wei, K.E., Rogers, R., DeGrasse, J.A., Chait, B.T., Aitchison, J.D., and Rout, M.P. (2007). Comprehensive analysis of diverse ribonucleoprotein complexes. Nat. Methods *4*, 951–956.

Penczek, P.A., Grassucci, R.A., and Frank, J. (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy *53*, 251–270.

Penczek, P.A., Frank, J., and Spahn, C.M.T. (2006). A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. J. Struct. Biol. *154*, 184–194.

Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*. J. Microsc. *146*, 113–136.

Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. *333*, 721–745.

Rubinstein, J.L., Walker, J.E., and Henderson, R. (2003). Structure of the mitochondrial ATP synthase by electron cryomicroscopy. EMBO J. *22*, 6182–6192.

Sander, B., Golas, M.M., and Stark, H. (2003a). Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra. J. Struct. Biol. *142*, 392–401.

Sander, B., Golas, M.M., and Stark, H. (2003b). Corrim-based alignment for improved speed in single-particle image processing. J. Struct. Biol. *143*, 219–228.

Sander, B., Golas, M.M., Makarov, E.M., Brahms, H., Kastner, B., Lührmann, R., and Stark, H. (2006). Organization of core spliceosomal components U5 snRNA loop I and U4/U6 di-snRNP within U4/U6.U5 tri-snRNP as revealed by electron cryomicroscopy. Mol. Cell *24*, 267–278.

Scheres, S.H.W., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P.B., Frank, J., and Carazo, J. (2007). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat. Methods *4*, 27–29.

Scheres, S.H.W., Melero, R., Valle, M., and Carazo, J. (2009a). Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. Structure *17*, 1563–1572.

Scheres, S.H.W., Valle, M., Grob, P., Nogales, E., and Carazo, J. (2009b). Maximum likelihood refinement of electron microscopy data with normalization errors. J. Struct. Biol. *166*, 234–240.

Schmid, M.F., and Booth, C.R. (2008). Methods for aligning and for averaging 3D volumes with missing data. J. Struct. Biol. *161*, 243–248.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H.D. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. Science *310*, 827–834.

Sigworth, F.J. (1998). A maximum-likelihood approach to single-particle image refinement. J. Struct. Biol. *122*, 328–339.

Simonetti, A., Marzi, S., Myasnikov, A.G., Fabbretti, A., Yusupov, M., Gualerzi, C.O., and Klaholz, B.P. (2008). Structure of the 30S translation initiation complex. Nature *455*, 416–420.

Sousa, D., and Grigorieff, N. (2007). Ab initio resolution measurement for single particle structures. J. Struct. Biol. *157*, 201–210.

van Heel, M. (1984). Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). Ultramicroscopy *13*, 165–183.

van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy *21*, 111–123.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J. Struct. Biol. *116*, 17–24.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

Zemlin, F. (1989). Dynamic focussing for recording images from tilted samples in small-spot scanning with a transmission electron microscope. J. Electron Microsc. Tech. *11*, 251–257.

